Niveau(x) d'enseignement: 4 Numéro EDUQ : 8670181
Centre d'intérêt: 570

Evaluation de trois modèles de tests diagnostiques en mathématique / par Richard G. Bertrand... [et al.]. -- Sainte-Foy : INRS-Education, 1984.
26 f. ; 28 cm.

Code : C-126. Bibliogr.: f. 24-26.

Vedette(s)-secondaire(s): Bertrand, Richard G., 1949- * INRS-Education

Descripteur(s): Evaluation de tests * Elaboration de tests * Test de diagnostic * Mathématiques * Secondaire (Ordre d'enseignement)

Existe-t-il un ou des modèles d'instruments diagnostiques qui favorisent mieux que d'autres l'éclosion de l'évaluation formative? L'objectif général de la recherche est d'évaluer trois modèles d'instruments diagnostiques en fonction d'une liste préétablie de cinq critères. Ces trois modèles sont: le modèle dit des erreurs caractéristiques, le modèle conçu à partir des objectifs d'apprentissage et le modèle à facettes. Quatre tests par modèle ont été élaborés. Les cinq critères d'évaluation sont: 1. la souplesse des instruments pour la gérance de l'enseignement; 2. la finesse des instruments pour détecter les préalables non maîtrisés; 3. la possibilité d'effectuer une analyse des patrons d'erreurs; 4. la pertinence des facettes; 5. l'unidimensionalité des catégories diagnostiques. Dix enseignants et plusieurs groupes d'élèves de la polyvalente de Tilly à Ste-Foy ont participé à l'expérimentation. Les tests ont été administrés entre mars et mai 1984. Les enseignants ont ensuite participé à une entrevue semi-dirigée. Les résultats des analyses et la considération des enseignants indiquent que le modèle à facettes est le meilleur instrument diagnostique.

Type de document: r
Notice(s) connexe(s): 8570816

EDUQ 8670181

EVALUATION DE TROIS MODELES DE TESTS DIAGNOSTIQUES EN MATHEMATIQUE.

par Richard G. Bertrand, Marcelle Tremblay-Desrochers, Monique Morin, Jocelyne Roberge-Brassard.

C-126

Octobre 1984

EDUQ 8670181

ÉVALUATION DE TROIS MODÈLES DE TESTS DIAGNOSTIQUES EN MATHÉMATIQUE

par

Richard G. Bertrand, Marcelle Tremblay-Desrochers
Monique Morin, Jocelyne Roberge-Brassard

C-126

INRS-Éducation
2383, Chemin Sainte-Foy
Sainte-Foy, Québec
G1V 1T1

Octobre 1984

INTRODUCTION

Cette recherche[1] vise à apporter une contribution à la conceptualisation de modèles d'outils diagnostiques aptes à améliorer l'évaluation de l'enseignement et de l'apprentissage de la mathématique au premier cycle de l'enseignement secondaire. Elle consiste à évaluer trois modèles d'instruments diagnostiques adaptés pour le secteur des nombres, à l'aide d'une liste de cinq critères. Le modèle qui répondra le mieux à nos critères sera utilisé par la suite pour étudier divers modes de gestion des instruments, tels la gestion par ordinateur personnel, la gestion par l'élève et la gestion par l'enseignant.

---

1. Cette recherche est financée par les fonds F.C.A.C. et C.R.S.H.

I RAISONS QUI ONT MOTIVÉ CETTE RECHERCHE

C'est à la suite de lectures et de réflexions portant sur le rôle et la pertinence de l'évaluation formative qu'est né le projet de recherche de modèles d'instruments diagnostiques en mathématique. Les dernières sessions de l'A.S.M.É. ont permis de nourrir plus particulièrement ces réflexions.

Plusieurs documents du ministère de l'Éducation soulignent l'importance de l'évaluation formative que ce soit la Politique générale d'évaluation pédagogique, l'École québécoise, ou le Perfectionnement. En outre, le Conseil supérieur de l'éducation stipule que le milieu scolaire constate des problèmes d'application de l'évaluation formative. Le Conseil recommande donc au ministère de développer des instruments de mesure à fonction formative à l'intention des enseignants.

Par ailleurs, bon nombre de chercheurs du Québec, des États-Unis et de l'Europe traitent des principaux aspects de l'évaluation formative: sa fonction diagnostique (Scallon, 1980; Bégin et Dussault, 1980), la possibilité accrue d'interaction entre les élèves et l'enseignant (Racine, 1982), le diagnostic "en cours de route" des faiblesses dues à l'enseignement ou à l'apprentissage (Popham, 1981), sa grande valeur pédagogique (Bloom, 1971), le rôle des hiérarchies d'apprentissage pour la construction d'instruments diagnostiques (Airasion, 1971), sa fonction de régulation interactive et individualisée (Allal, Cardinet et Perrenoud, 1980), et enfin son rôle de retour d'information permettant d'adapter l'enseignement (Cardinet, 1982).

Certains chercheurs ont même publié des résultats d'études empiriques qui révèlent que

- l'enseignement correctif devrait se baser sur les préalables plutôt que sur les notions testées (Fiel et Okey, 1975);
- les items d'un test diagnostique devraient être intimement liés au contenu enseigné (Kidd et alii., 1970);
- pour construire un test diagnostique on pouvait employer ce qu'on appelle des catégories diagnostiques (Dassa, 1982), ou encore la théorie des facettes (Scallon, 1981).

Dans le domaine de l'évaluation formative, en mathématique plus particulièrement, James Sherril écrit qu'il y a peu de secteurs de l'enseignement de la mathématique aussi mûrs pour la recherche que le diagnostic pédagogique et l'enseignement correctif.

Ces lectures et ces réflexions nous ont mené à considérer deux grandes approches en évaluation formative: l'approche clinique et l'approche métrique. L'approche clinique est celle qui est caractérisée plutôt par l'utilisation de schémas individuels d'entrevue et des protocoles d'observation systématique ("checklist") de l'élève. Cette approche a connu un certain essor depuis quelques années grâce à la mise sur pied de cliniques mathématiques indépendantes de la classe et donc de l'enseignant.

L'approche métrique est basée surtout sur l'utilisation d'instruments de mesure diagnostiques en classe même. C'est donc l'enseignant qui en est le responsable.

Notre recherche se situe dans le contexte de cette deuxième approche. Bien sûr, certains élèves auront toujours besoin d'être observés et interviewés. Mais ces observations et ces entrevues sont onéreuses, qu'elles se situent en classe ou dans une clinique mathématique. Nous sommes d'avis qu'il est possible de réduire au minimum ces entrevues et ces observations par un travail visant à augmenter autant la précision

des tests diagnostiques que leur degré d'utilisation par les enseignants à l'aide d'un mode de gestion adéquat. De cette façon, nous pensons qu'il est possible de diagnostiquer bon nombre de difficultés de la part des élèves, qu'elles soient dues à des erreurs simples de calcul ou à des fausses procédures plus difficiles à détecter.

Cependant le diagnostic de ces difficultés n'est possible que grâce à une adéquation des instruments à l'enseignement. C'est pourquoi un soin très particulier doit être accordé à la facture des instruments diagnostiques.

Ces considérations nous ont amenés à poser le problème général de la présente étude en ces termes: existe-t-il un ou des modèles d'instruments diagnostiques qui favorise(nt) mieux que d'autres l'éclosion de l'évaluation formative?

II OBJECTIFS DE LA RECHERCHE ET CRITÈRES D'ÉVALUATION

II.1 Objectifs

La problématique qui sous-tend ce projet nous a conduit à énoncer l'objectif général de la recherche de la façon suivante:

- évaluer trois modèles d'instruments diagnostiques en fonction d'une liste préétablie de cinq critères.

Deux sous-objectifs précisent cet objectif général:

- identifier les forces et les faiblesses de chacun des trois modèles en tant que générateurs d'outils diagnostiques;
- juger de la pertinence d'application des trois modèles pour le secteur des nombres en mathématiques.

II.2 Critères d'évaluation

Le problème général de l'étude énoncé plus haut peut facilement nous entraîner à vouloir étreindre une réalité fort complexe. C'est pourquoi nous avons senti le besoin de spécifier une liste de critères suffisamment précis pour que l'évaluation des modèles soit la plus opérationnelle possible. Cette liste de critères reflète ce que nous entendons exiger des instruments de mesure diagnostiques.

Trois de ces critères sont d'ordre pratique:

1) la souplesse des instruments pour la gérance de l'enseignement qui fait référence à la facilité d'utilisation des instruments en classe. C'est vraiment l'adéquation du matériel à l'enseignement qui nous intéresse ici.
2) la finesse des instruments pour détecter les préalables non maîtrisés que nous jugeons capitale pour la planification de l'enseignement correctif. Nous voulons savoir dans ce cas si le modèle permet un diagnostic basé sur une hiérarchie des notions préalables.
3) la possibilité d'effectuer une analyse des patrons d'erreurs qui sont si fréquents lors d'opérations mathématiques sur les nombres.

Les deux autres critères sont d'ordre méthodologique:

4) la pertinence des facettes qui se fait grâce à l'analyse des composantes de la variance et dont il sera question plus loin.
5) l'unidimensionalité des catégories diagnostiques qui consiste à vérifier si les items se regroupent de façon homogène en une catégorie diagnostique d'erreurs définie préalablement. L'analyse factorielle est employée à cette fin.

III DESCRIPTION DES MODÈLES

a. Leur origine

Les trois modèles d'instruments diagnostiques retenus pour l'évaluation sont: le modèle dit des erreurs caractéristiques (Modèle I), le modèle conçu à partir d'objectifs d'apprentissages (Modèle II) et le modèle à facettes (Modèle III).

Des tests ont été élaborés en rapport avec chacun de ces modèles mais avant d'y arriver, nous avons recueilli de l'information sur les difficultés les plus souvent rencontrées dans les opérations mathématiques préalables à l'apprentissage de l'algèbre.

Cette cueillette d'informations s'est faite à partir de témoignages d'enseignants ainsi que de recherches empiriques qui se sont déroulées dans le milieu scolaire (Projet: Étude de la classe et de son environnement de l'INRS-Éducation, tests du centre MEDIC de l'Université de la Colombie-Britannique).

Les difficultés relevées ont ensuite été catégorisées (ex: ppcm, simplification de fractions, arrondissement des nombres à virgule...) puis apparentées à l'un ou l'autre des quatre secteurs de nombres suivants: nombres naturels, nombres entiers, fractions, nombres à virgule.

Nous avions donc un contenu spécifique d'évaluation que nous allions retrouver d'un modèle à l'autre. Finalement, quatre tests par modèle ont été élaborés soit un test pour chacun des secteurs; ce qui totalise douze tests (voir Tableau III.1).

Tableau III.1

SECTEURS DES NOMBRES
COUVERTS PAR LES TROIS MODÈLES

MODÈLE I (erreurs caractéristiques)	MODELE II (objectifs)	MODELE III (facettes)
I.1 Nombres naturels	II.1 Nombres naturels	III.1 Nombres naturels
I.2 Nombres entiers	II.2 Nombres entiers	III.2 Nombres entiers
I.3 Fractions	II.3 Fractions	III.3 Fractions
I.4 Nombres à virgule	II.4 Nombres à virgule	III.4 Nombres à virgule

b. Le modèle I (ou modèle des erreurs caractéristiques)

Le modèle I s'inspire d'un modèle que M. Clément Dassa a développé à l'Université de Montréal et qui est essentiellement basé sur la notion d'erreur caractéristique. Une erreur caractéristique est une erreur précise survenant de façon fréquente chez les élèves. Dans notre recherche, les erreurs caractéristiques proviennent des catégories de difficultés que nous avions préalablement obtenues. Par exemple, pour une catégorie de difficultés donnée comme "division de deux nombres naturels", nous avons identifié l'erreur caractéristique suivante: "complète mal la division (le reste est supérieur au diviseur)". Nous avons procédé de la même façon pour les autres catégories de difficultés afin d'obtenir une liste des erreurs caractéristiques pour chacun des secteurs du modèle I (voir Tableau III.2).

Dans l'élaboration des tests de ce modèle, deux items par erreur caractéristique ont été construits. Il est important de préciser qu'un

Tableau III.2

Liste des erreurs caractéristiques et numéros des items associés.
(Modèle I)

I.1 Nombres naturels

No d'items	Erreurs caractéristiques
1, 2	Complète mal la division (le reste est supérieur au diviseur).
3, 4	Mélange le plus petit commun multiple (p.p.c.m.) avec le plus petit commun diviseur différent de 1.
5, 6	Multiplie la base par l'exposant.
7, 8	Confond l'ordre de priorité des opérations.
9, 10	Oublie d'abaisser le zéro au dividende.
11, 12	Ne distingue pas "facteurs" de "facteurs premiers".
13, 14	Mélange les symboles $<$, $>$, $\geq$, $\leq$.

item ne détecte qu'une et une seule de ces erreurs. Pour répondre à un item, l'élève doit solutionner deux exercices et encercler la lettre correspondant à la fois aux deux bonnes réponses (voir Tableau III.3).

Lors de la passation d'un des tests, l'enseignant était libre d'administrer le test entier ou seulement certains items afin de s'assurer de la présence des erreurs caractéristiques.

c. Le modèle II (ou modèle centré sur des objectifs)

Le modèle II a été conçu à partir d'objectifs spécifiques d'apprentissage correspondant aux catégories de difficultés déjà identifiées. Ces objectifs ainsi que les items qui en découlent proviennent d'un modèle réalisé par Michel Labelle et intitulé "BATMANS-1" (1980). Ce modèle contient des objectifs d'apprentissage et une banque d'items dans le domaine de la mathématique au premier cycle du secondaire. Nous avons donc apparenté un ou des objectifs spécifiques aux catégories de difficultés préalablement définies pour ainsi obtenir une liste d'objectifs d'apprentissage pour chacun des secteurs de nombre déjà identifié (voir Tableau III.4). Un à deux items ont été associés à chacun de ces objectifs (voir Tableau III.5). L'enseignant était libre d'administrer un test en entier ou seulement certains items reliés à des objectifs précis.

d. Le modèle III (dit à facettes)

Le modèle III a été conçu par le professeur Gérard Scallon sur une idée empruntée à la théorie des facettes de Guttman (1954).

Pour ce modèle, les catégories de difficultés s'inscrivent dans des tâches mathématiques précises. On forme ensuite un tableau nommé schéma

Tableau III.3

I.1 Nombres naturels

No d'items	Erreur caractéristique
1, 2	Complète mal la division (le reste est supérieur au diviseur).

Item 1 . Effectue les divisions suivantes.

Exercice 1 : 357 ÷ 14 Exercice 2 : 588 ÷ 13

a) ex. 1 : 24 reste 21 et ex. 2 : 45 reste 3
b) ex. 1 : 24 reste 21 et ex. 2 : 44 reste 16
c) ex. 1 : 25 reste 7 et ex. 2 : 44 reste 16
d) ex. 1 : 25 reste 7 et ex. 2 : 45 reste 3

Item 2 . Effectue les divisions suivantes.

Exercice 1 : 496 ÷ 17 Exercice 2 : 627 ÷ 13

a) ex. 1 : 28 reste 20 et ex. 2 : 47 reste 16
b) ex. 1 : 29 reste 3 et ex. 2 : 48 reste 3
c) ex. 1 : 28 reste 20 et ex. 2 : 48 reste 3
d) ex. 1 : 29 reste 3 et ex. 2 : 47 reste 16

Tableau III.4

Liste des objectifs et numéros des items associés.
(Modèle II)

II.1 Nombres naturels

No d'items	Objectifs
1, 2	Diviser deux nombres dont le diviseur est inférieur à 1000 dans $\mathbb{N}$.
3, 4	Décrire en extension l'ensemble des facteurs premiers d'un nombre naturel.
5	Décrire en extension l'ensemble des multiples d'un nombre naturel.
6	Calculer le plus petit commun multiple de 2 ou 3 nombres naturels.
7, 9	Écrire une puissance sous forme de facteurs identiques.
8	Écrire un produit de facteurs identiques sous forme de puissance.
10	Utiliser les symboles $>$, $\not>$, $<$, $\not<$, dans la comparaison de nombres naturels.
11	Effectuer des calculs en respectant l'ordre des opérations dans $\mathbb{N}$.

II.1 Nombres naturels

No d'items	Objectif
1, 2	Diviser deux nombres dont le diviseur est inférieur à 1000 dans N.

Item 1 . Effectue : 531 378 ÷ 477 =

a) 10 113
b) 1 013
c) 1 113
d) 1 114

Item 2 . Effectue : 308 025 ÷ 555 =

a) 555
b) 5055
c) 5550
d) 55

à facettes lequel comprend des opérations mathématiques hiérarchisées où les difficultés inventoriées sont susceptibles d'être rencontrées.

Voici trois schémas à facettes correspondant à trois tâches précises (voir Tableau III.6):

- Division de deux nombres naturels.
- Addition de deux nombres à une puissance.
- Calculer la puissance d'un nombre.

Le premier schéma correspondant à la tâche: "division de deux nombres naturels" est formé d'une seule facette qui comprend trois éléments.

Le deuxième schéma est formé, quant à lui, d'une seule facette à deux éléments.

Il arrive que le schéma contienne plus d'une facette. Alors les éléments des facettes seront croisés entre eux. Dans le troisième schéma, la première facette (F) indique l'expression de la puissance et la deuxième (G) les types de nombres. L'item correspondant au numéro 1.11 porte sur la multiplication de nombres naturels alors qu'à l'item 3.21 on doit calculer la puissance de nombres rationnels. Les numéros 1.11 et 3.21 et tous les autres inscrits dans chaque cellule du schéma à facettes correspondent à des items que l'on retrouve dans l'un ou l'autre des tests.

Prenons un exemple pour illustrer ce qui précède.

On a trouvé que plusieurs élèves éprouvaient de la difficulté à diviser deux nombres naturels. La division de deux nombres naturels constitue donc <u>la tâche</u> qui nous intéresse. Nous avons alors défini trois situations (ou <u>éléments</u>) qui permettront d'établir un diagnostic valable après la passation du test. Voici comment s'inscrivent ces données:

Tableau III.6

TACHE : DIVISION DE DEUX NOMBRES NATURELS (F.D. 1.1)

EXPRESSION DU QUOTIENT

SANS RESTE	AVEC RESTE	AVEC RESTE EXPRIME SOUS FORME DECIMALE
1.1	1.3	1.5
1.2	1.4	1.6

TACHE : ADDITION DE DEUX NOMBRES A UNE PUISSANCE (F.D. 1.2 + F.D. 3.2)

TYPES DE NOMBRES

NOMBRES NATURELS	NOMBRES RATIONNELS
1.9	3.9
1.10	3.10

TACHE : CALCULER LA PUISSANCE D'UN NOMBRE (F.D. 1.3 + F.D. 3.6)

EXPRESSION DE LA PUISSANCE (F)

TYPES DE NOMBRES (G)	MULTIPLICATION	CALCUL DE LA PUISSANCE	ECRITURE D'UN NOMBRE A UNE PUISSANCE DONNEE
NOMBRES NATURELS	1.11	1.15	1.13
	1.12	1.16	1.14
NOMBRES RATIONNELS	3.17	3.21	3.19
	3.18	3.22	3.20

Tâche: Division de deux nombres naturels.
La facette: Expression du quotient.
Éléments de la facette:

1. sans reste
2. avec reste
3. avec reste exprimé sous forme décimale.

Par la suite, nous plaçons ces divers points dans le schéma à facettes (F.D.1.1) que l'on retrouve au haut du Tableau III.6.

Ces quelques informations constituent, pour l'essentiel, la description du modèle III dit à facettes.

Son utilisation

Lors de l'expérimentation, les enseignants avaient à choisir entre deux possibilités:

1. Ils pouvaient administrer en une seule fois tous les items d'un des quatre secteurs dont nous avons parlé préalablement.
2. Ils pouvaient aussi choisir de n'administrer que les items se rapportant à un seul schéma à facette à la condition que la tâche décrive le domaine qui doit être investigué. La reproduction de la tâche "division de deux nombres naturels" et du schéma à une seule facette "expression du quotient" (F.D.1.1.) que l'on retrouve au Tableau III.7 illustrent l'origine des six items qui se retrouvent dans le test. L'enseignant qui désirait vérifier le degré de maîtrise de cette tâche devait administrer le sous-test comportant ces six items.

Tableau III.7

TACHE : DIVISION DE DEUX NOMBRES NATURELS (F.D. 1.1)

EXPRESSION DU QUOTIENT

SANS RESTE	AVEC RESTE	AVEC RESTE EXPRIME SOUS FORME DECIMALE
1.1	1.3	1.5
1.2	1.4	1.6

EFFECTUE LES DIVISIONS SUIVANTES.

Item 1 . 784 ÷ 28 Réponse

Item 2 . 476 ÷ 17 Réponse

EFFECTUE LES DIVISIONS SUIVANTES EN INDIQUANT LE RESTE.

Item 3 . 3213 ÷ 15 Réponse

Item 4 . 4123 ÷ 14 Réponse

EFFECTUE LES DIVISIONS SUIVANTES EN EXPRIMANT LE RESTE SOUS FORME DÉCIMALE.

Item 5 . 2317 ÷ 25 Réponse

Item 6 . 5835 ÷ 12 Réponse

IV EXPÉRIMENTATION

Quatre enseignants de deuxième et de troisième secondaire de la polyvalente Samuel de Champlain de Beauport, en banlieue de Québec, ont consenti à la pré-expérimentation de nos tests.

Leurs 280 élèves ont été répartis en trois groupes. À chacun des groupes, on a administré respectivement les tests des quatre secteurs d'un des trois modèles.

Suite à cette pré-expérimentation et à la rencontre que nous avons eue avec les enseignants, des modifications, ajouts ou retraits furent effectués à nos instruments avant d'entreprendre la phase expérimentale.

Celle-ci a eu lieu dans une autre polyvalente de la banlieue, la régionale de Tilly à Ste-Foy.

Deux groupes d'élèves de deuxième secondaire ont utilisé tous les tests du modèle I, deux autres groupes ceux du modèle II et enfin les deux derniers groupes ceux du modèle III. Les données recueillies lors de cette administration ont permis d'évaluer les modèles en fonction des critères 4 et 5 d'ordre méthodologique. Puis plusieurs groupes, distincts des six précédents, ont utilisé quelques tests de l'un ou l'autre modèle, en tout ou en partie, au gré des dix enseignants participants. Ceux-ci ont administré les tests entre les mois de mars et mai 1984. Finalement, chacun des dix enseignants a été soumis à une entrevue semi-dirigée portant sur l'appréciation des modèles en fonction des critères 1, 2 et 3 d'ordre pratique dans le but de:

- juger de la souplesse des instruments en terme d'intégration à l'enseignement (critère 1),
- voir si ces tests peuvent effectivement détecter les préalables à l'algèbre non-maîtrisés (critère 2),

- déceler une erreur qui revient d'une manière systématique (critère 3).

V RÉSULTATS

a. Statistiques

Les trois modèles de notre étude ont fait l'objet d'analyses statistiques afin de voir comment chacun se comportait.

Une méthode bien précise a été employée répondant aux caractéristiques propres de chacun des modèles.

Modèle I

Pour le modèle I (modèle des erreurs caractéristiques), chacun des quatre tests a été analysé au moyen d'une analyse factorielle (J-factor, SPSS). Le but de cette analyse est d'identifier des catégories diagnostiques et voir si les items se regroupent bien dans les différentes catégories diagnostiques préalablement identifiées.

En général, nous avons obtenu peu d'informations pratiques et utilisables sur les structures des catégories diagnostiques. Tout au plus l'analyse factorielle confirme-t-elle certaines structures que nous avions définies à priori, nous assurant ainsi de l'unidimensionalité des catégories diagnostiques (critère 5).

Par exemple, l'analyse montre que deux items que nous jugions équivalents se regroupent autour d'un même facteur et il est rare d'en voir plus de deux se regrouper ensemble. Donc, les deux items associés à une erreur caractéristique formaient un tout identifiant une catégorie diagnostique distincte.

Modèle II

Dans le cas du modèle II (modèle des objectifs), nous avons procédé à une analyse des leurres pour chacun des items des quatre tests, chaque item ayant quatre choix de réponses. Les résultats révèlent que très peu d'items ont montré une grande valeur diagnostique. Plusieurs de ceux-ci ont très bien été réussis, ce qui indique que les leurres n'ont pas été choisis souvent. De même, pour les items qui ne sont pas réussis par plus de 65% des élèves, en général un seul leurre parfois deux sont choisis par plus de 10% des élèves. Ainsi très peu de leurres jouent un rôle vraiment important confirmant par le fait même la faiblesse diagnostique des items de ce modèle.

Modèle III

Pour le modèle III (modèle à facettes), l'analyse des composantes de la variance, au sens de la théorie de la généralisabilité, a servi pour l'étude des tableaux à facettes (critère 4). Plusieurs de ces tableaux offrent un intérêt sur le plan diagnostique. Par exemple, le Tableau V.1 portant sur l'équivalence entre un nombre à virgule et un pourcentage est intéressant à plus d'un point de vue. Tout d'abord les items à l'intérieur d'une même cellule ont une moyenne relativement semblable ce que reflète le faible 7.3% d'erreur. De plus, on note une variation systématique des moyennes des items entre les trois niveaux "2", "1" et "3" de la facette qui est reflété par le pourcentage (12%) de la variance de la facette F. Cependant, le pourcentage de variance le plus important est assurément celui de l'interaction FXS entre la facette et les sujets (plus de 40%) indiquant que les élèves ont un profil de réponses très différent d'un niveau à l'autre.

Tous les tableaux à facettes n'ont toutefois pas autant de valeur diagnostique. Plusieurs en effet comportent des items réussis à plus de

Tableau V.1

TACHE : EQUIVALENCE ENTRE NOMBRE A VIRGULE ET POURCENT (F.D. 4.5)

NOMBRE DE CHIFFRES SIGNIFICATIFS APRES LA VIRGULE

2	1	3
4.19	4.17	4.21
4,20	4.18	4.22

SOURCE	SOMME DES CARRES	DEGRE DE LIBERTE	CARRE MOYEN	COMPOSANTE DE VARIANCE	%
S	27.7960	57	.4876	.07888	40.082
F	9.8161	2	4.9080	.02367	12.027
F X S	19.8506	114	.1741	.07988	40.590
ERREUR	2.5000	174	.0144	.01437	7.301
TOTAL	59.9626	347		.19680	

TACHE : EQUIVALENCE ENTRE NOMBRE A VIRGULE ET POURCENT (F.D. 4.5)

VARIABLE	4.19	4.17	4.21
MOYENNE	.93	.86	.57
ECART-TYPE	.26	.35	.50
VARIABLE	4.20	4.18	4.22
MOYENNE	.91	.88	.52
ECART-TYPE	.28	.33	.50

TAILLE : 58

80% ce qui montre que la facette à un faible pourcentage de variation. Par contre, l'interaction avec les sujets donne souvent un pourcentage de variance assez important (10 à 50%). Certains tableaux par ailleurs révèlent que ni la facette ni l'interaction de la facette avec les sujets ne donne un pourcentage de variance important: les pourcentages n'atteignent même pas 10%.

En général, le modèle nous apparaît tout de même très riche en informations diagnostiques.

b. Les entrevues avec les enseignants

Au terme de cette recherche, nous avons voulu rencontrer les dix enseignants ayant participé à l'expérimentation afin d'obtenir leurs commentaires sur des points précis concernant les trois modèles des tests diagnostiques. Le protocole d'entrevue élaboré à cette fin comprenait les huit points suivants sur lesquels ils devaient se prononcer.

1. Les raisons préalables au choix des modèles et secteurs.
2. La distinction des modèles au niveau de l'utilisation en classe.
3. Les modèles appropriés pour:
   a) la détection des préalables à l'algèbre non-maîtrisés
   b) la mise en évidence d'une erreur se produisant systématiquement
   c) la planification d'un enseignement correctif.
4. La souplesse du modèle par rapport à l'enseignement.
5. Les interventions subséquentes auprès des élèves.
6. La correspondance entre ces tests et les attentes en regard d'un test diagnostique.
7. Le niveau le plus pertinent pour l'administration de ces tests et le temps de l'année le plus approprié.
8. L'avenir du diagnostic pédagogique.

De ces entrevues, il ressort que les enseignants évoquent souvent le facteur temps et la longueur du test pour justifier le choix du modèle qu'ils ont fait. Bien peu d'entre eux optent pour les items à choix multiples. On préfère des tests diagnostiques dont les items sont sans choix de réponse tel que proposé dans le modèle des facettes. Celui-ci est d'ailleurs le plus apprécié des enseignants. On reconnaît, en effet qu'il demande un plus grand effort à l'élève qui doit effectuer toutes les opérations avant de trouver la réponse à chacun des items. Pour les enseignants cette démarche permet une plus grande précision puisque l'on peut, de cette façon mettre le doigt plus facilement sur la difficulté. Il a été souligné aussi que la hiérarchie des difficultés assure une meilleure structure des items. Ce modèle permet donc de mieux détecter les difficultés individuelles. Puisqu'il ressemble à bien des égards au type d'exercices utilisés par les enseignants, il s'intègre à leur enseignement de façon naturelle. Selon eux, un test diagnostique doit donner des informations précises à toutes les étapes d'une opération mathématique et le modèle des facettes fournit ce type d'informations. Pour toutes ces raisons, le modèle des facettes est celui qui, selon les enseignants, permet le mieux de détecter les préalables à l'algèbre dans le secteur des nombres qui n'ont pas été maîtrisés par les élèves; de déceler une erreur qui revient de façon systématique et de planifier un enseignement correctif. Il est à noter que la presque totalité des enseignants sont intervenus après la passation des tests diagnostiques, soit par une correction avec le groupe, quelques explications collectives ou, dans quelques cas, par une remédiation auprès des élèves pris individuellement. Toutes ces considérations de la part des enseignants et les résultats découlant de la recherche sur l'évaluation des modèles justifient, selon nous, le choix du modèle des facettes dans l'étude que nous projetons d'effectuer sur les modes de gestion.

c. Commentaires des élèves

Lors de la passation des tests, les élèves furent invités à répondre aux deux questions suivantes.

- As-tu aimé ce test?
- Penses-tu que cela va te rendre service dans ce que tu apprends?

Ils pouvaient émettre aussi leurs commentaires personnels. Ce questionnaire avait pour but de voir quelle attitude ils ont face à des tests comme ceux-là.

Plusieurs élèves ont souligné l'intérêt de ces tests parce qu 'ils leur permettent de réviser leurs connaissances en mathématique. D'autres avouent les trouver ou trop faciles ou trop difficiles selon le degré de maîtrise qu'ils en ont. On critique la longueur de certains tests surtout dans le secteur des fractions. Les élèves ont vu le modèle I comme un moyen de faire de la révision. Le modèle II suscite peu de commentaires parce que peu utilisé dans l'expérimentation. Quant au modèle III, il permet, semble-t-il, aux élèves de voir où sont leurs erreurs.

Enfin, on semble reconnaître que le secteur 4, celui des nombres à virgule, identifie davantage les difficultés spécifiques.

VI PROSPECTIVES

Suite à cette recherche, nous constatons qu'un travail important reste à faire au niveau du mode de gestion de ces instruments diagnostiques. Plusieurs questions surgissent à ce sujet. De quelle façon les tests diagnostiques doivent-ils être construits? Comment et à qui administrer ces instruments? Qui doit les corriger? De quelle façon les

résultats des élèves seront-ils compilés puis analysés? Quel mode d'intervention suivra ces tests?

Ces questions nous préoccupent et feront l'objet d'une prochaine recherche dans laquelle nous nous pencherons sur le problème suivant: existe-t-il un mode de gestion des instruments diagnostiques qui peut davantage rendre viable et rentable la pratique de l'évaluation formative en classe?

BIBLIOGRAPHIE

AIRASIAN, P.W., (1971) The role of evaluation in mastery learning, in Mastery learning : theory and practice, J.H. Block (éd.), H.R.W. INC., New-York.

ALLAL, L., (1980) Recherche sur les interactions aptitude-traitement : implications pour l'évaluation formative, Éducation et Recherche, Vol. 3, Genève, Suisse.

ÉGIN, Y., DUSSAULT, G., (1980) La pédagogie de la réussite ou la redécouverte du bon enseignement, Vie pédagogique, Vol. 10, Ministère de l'éducation du Québec, Montréal.

BERTRAND, R., TREMBLAY-DESROCHERS, M., MORIN, M., ROBERGE-BRASSARD, J., Guides (3) d'utilisation des tests diagnostiques pour les modèles I, II, III, Documents D-172, D-173, D-174, INRS-Éducation, Québec, février 1984.

BERTRAND, R., TREMBLAY-DESROCHERS, M., MORIN, M., ROBERGE-BRASSARD, J., Tests (12) diagnostiques pour les modèles I, II, III, Documents T-50 à T-61, INRS-Éducation, Québec, février 1984.

BERTRAND, R., (1979) Évaluation d'une calculatrice auto-correctrice : Classmate 88, Revue des sciences de l'éducation, Vol V, N° 1.

BLOOM, B.S., (1971) Mastery learning, in Mastery Learning : theory and practice, J.H. Block (éd.), H.R.W. INC., New-York.

BOILY, C., OUELLETTE et alii (1980) Questionnaire aux étudiants I, Document INRS-Éducation, T-42, Québec.

BROWN, J.S., BURTON, R.R., (1978) Diagnostic models for procedural bugs in basic mathematical skills, Cognitive Science, Vol. 2.

CARDINET, J., (1982) L'évaluation formative à l'école primaire, Education et Recherche, Vol. 3.

CHARBONNEAU, B., et alii (1984) L'ordinateur, outil de gestion pédagogique d'une classe, Les cahiers d'enseignement et de recherche en technologie éducative, N° 4, Université du Québec à Hull, Hull.

Conseil supérieur de l'éducation (1982) L'évaluation des apprentissages : "ça comptes-tu ?", Québec.

DASSA, C., (1982) L'identification des catégories diagnostiques d'une batterie de tests diagnostiques, Mesure et évaluation en éducation, Vol 5, N° 5, Québec.

DE GRUIJTER, D.N.M., (1980) A two-stage testing procedure, in Psychometrics for Educational Debates, L.J.Th. Van der Kamp, W.F. Langerak, D.N.M. de Gruijter (éds), John Wiley & Sons Ltd., Chichester.

FERRARIS, M., MIBORO, V., OLIMPO, G., (1983) Diagnostic testing and the development of CAL remedial sequences, Comptes-rendus du 4e Symposium canadien sur la technologie pédagogique, Winnipeg, Canada.

FIEL,R.L., OKEY, J.R., (1975) The effects of formative evaluation and remediation on mastery of intelligence skills, The Journal of Educational Research, Vol 68, N° 7.

FISCHER, G.H., PENDL, P., (1980) Individualized testing on the basis of the dichotomou Rasch model, in Psychometrics for Educational Debates, L.J.Th. Van der Kamp, W.F. Langerak, D.N.M. de Gruijter (éds), John Wiley & Sons Ltd, Chichester.

Gouvernement du Québec, Ministère de l'éducation (1981), Politique générale d'évaluation pédagogique, Direction de l'évaluation pédagogique, D.G.D.P., document 16-0064, Québec.

Gouvernement du Québec, Ministère de l'éducation (1979) L'École québécoise, Énoncé de politique et plan d'action, Québec.

GREEN, B.F., (1983) Adaptive testing by computer, in Measurement, Technology, and Individuality in Education, R.B. Ekstrom (éd.), New Directions for Testing and Measure ment, N° 17, San Francisco.

KIRK, R.E. (1982) Experimental design, second edition, Wadsworth, INC., Belmont, California.

LABELLE, M., (1981) Batterie de tests de mathématique au niveau secondaire I (BATMANS-I), document d'information, Direction de l'évaluation pédagogique, D.G.D.P., Ministère de l'éducation du Québec, Québec.

LORD, F.M., (1970) Some test theory for tailored testing, in Computer-assisted Instruction, Testing, and Guidance, W. Holtzman (éd.), Harper & Row, New-York.

LORD, F.M., (1971) A theoretical study of two-stage testing, Psychometrika, Vol 36.

LORD, F.M., (1980) Some how and which for practical tailored testing, in Psychometrics for Educational Debates, L.J.Th. Van der Kamp, W.F. Langerak, D.N.M. de Gruijter (éds) John Wiley & Sons Ltd, Chichester.

MCARTHUR, D., et alii (1982) Diagnostic testing system : a complete diagnostic multiple-choice test package for the Apple II, ERIC Reports, EDRS, Washington, D.C.

MITCHELL, A.C., (1982) Using microcomputers to help teachers to develop their assessment procedures : a development project report, Programmed Learning and Educational Technology, Vol 19, N° 3.

MUIZNIEKS, V. DENNIS, J.R., (1979) A look at computer-assisted testing operations. The Illinois Series on educational application of computers, N° 12e., ERIC Reports, EDRS, Washington, D.C.

POPHAM, W.J., (1981) Modern educational measurement, Prentice-Hall Inc., Englewood Cliffs, N.J.

RACINE, S.B., (1982) L'évaluation formative : diagnostic et prescription pédagogique, Communication présentée à la 5e Session d'études de l'Association des spécialistes de la mesure et de l'évaluation en éducation, Québec.

SAMEJINA, F., (1977) The use of the information function in tailored testing, Applied Psychological Measurement, Vol. 1.

SCALLON, G., (1980) L'évaluation formative : un outil au service de l'enseignement et de l'apprentissage, document miméographié, Commission scolaire des Milles-Îles, Laval, P.Q.

SCALLON, G., (1981) La construction d'un test diagnostique selon des facettes, Partie I : présentation d'un modèle de recherche pédagogique, Monographies en mesure et évaluation, Département de mesure et évaluation, Université Laval, Québec.

SCALLON, G., (1984) Étude de faisabilité d'une démarche d'auto-diagnostic, rapport de recherche (à paraître), Université Laval, Québec.

SIGNER, B., (1982) Math doctor, M.D.-Microcomputer adaptive diagnosis, The Computing Teacher, Vol 10, N° 4.

TRAUB, R.E., WOLFE, R.G., (1981) Latent trait theories and the assessment of educational achievement, Review of Research in Education, Vol 9.

TRAVIS, B., (1984) Computer diagnosis of algorithmic errors, in Computer in Mathematics Education, V.P.Hansen & M.J. Zweng (éds), National Council of Teachers of Mathematics, INC., Yearbook.